**Predicting Concrete Strength with Machine Learning: Analysis of Linear Regression, K-Nearest Neighbours, Decision Trees, and Random Forest Models**.

Pankaj Shahare([2021bit058@sggs.ac.in](mailto:2021bit058@sggs.ac.in)), Dr. A.D. Sawarkar([adsawarkar@sggs.ac.in](mailto:adsawarkar@sggs.ac.in))

1. Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Department of Information Technology.
2. Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Department of Information Technology.

**Abstract:**

In the construction industry machine learning is employed to forecast the strength of substances such as concrete. In the future, more advanced algorithms will help predict how buildings and structures will carry out in accordance with the materials used and production processes. By using artificial intelligence, we can enhance concrete production by developing specialized algorithms to analyse its properties. Our study focused on comparing Four machine learning methods—**Multiple Regression, K-Nearest Neighbours, Decision Tree, Random Forest**— to forecast the strength of concrete. We used a large dataset we've gathered over time. The goal is to improve construction processes by accurately predicting concrete strength. Our research found that artificial intelligence methods, particularly those mentioned, can effectively determine the compressive strength of concrete that self-compacts. In simpler terms, we're using fancy computer techniques to figure out how strong concrete will be. By comparing different methods, we've found ways to predict concrete strength better. This helps us make better-quality concrete for building things. The models we created had an average error rate between 6.15% and 7.89%. They can be easily used in making building materials and ensuring their quality without needing a lot of computing power.

**Keyword:**

Concrete Strength Prediction, Machine Learning, Regression Models, K-Nearest Neighbour, Decision Tree, Random Forest, Building Materials

1. **Introduction:**

The construction industry stands as a cornerstone of economic development worldwide. Its significance is underscored by the increasing demand for buildings, infrastructure, and urbanization initiatives in burgeoning cities and densely populated regions. However, within this expansive sector lies the intricate process of producing building materials—an essential element bridging manufacturing and construction. Consider concrete, the backbone of construction projects globally. Concrete embodies a dual identity, simultaneously representing both the manufacturing realm and the construction domain. It serves as a pivotal material in factory production and finds application in on-site construction activities such as monolithic concreting. Despite its ubiquitous presence, predicting the characteristics of concrete provide a difficult task. This challenge arises from the intricate nature of concrete, one of the most complex artificial composites devised by humankind. Numerous variables influence the final quality of concrete, ultimately impacting the safety and reliability of structures, buildings, and infrastructure constructed from it [1]. Engineers and scientists in the field of materials science are tasked with the monumental responsibility of refining concrete production methods. Their objective is to achieve precise control over concrete's structure and properties, ensuring optimal performance in diverse applications. However, this pursuit is fraught with hurdles, particularly concerning human intervention [1].

Human error, whether in calculation, recipe formulation, or technological implementation, poses a significant risk to construction endeavours. Mistakes made by technologists can lead to catastrophic consequences, including construction accidents, structural failures, and premature collapse of load-bearing elements. Moreover, varying types of concrete structures are also susceptible to the repercussions of human fallibility. Considering these challenges, the construction industry stands at the precipice of digital transformation. Traditional notions of construction processes are being upended as digitalization ushers in a new era of innovation and efficiency [2]. However, compared to other industries like mechanical engineering, the construction sector has lagged in the adoption of new information technologies.

However, given the necessity to increase production, efficiency, and competitiveness, it is hard to evade the path towards digitization [3]. Any business that fails to adopt such a change risk becoming irrelevant in the current market where firms with advanced technology have an added advantage. In this journey of transformation, Artificial intelligence is among the most crucial things since it offers unparalleled opportunities for both creativity and simplification. These firms can totally alter how things are done at any point in the project's life cycle by utilizing a variety of data science technologies, such as big data analytics. Decision making will be improved at every level starting from design through execution up-to quality control thanks to AI technologies. Still, manufactured goods’ quality will also benefit from these systems while service delivery gets optimized in addition fostering innovation within construction industry setup. When we talk about adopting construction practices based on Artificial Intelligence methods, we mean that there should be shift in thinking pattern from being rigidly bound by historical information towards using flexible systems driven by facts generated through various technological processes, increasing their adaptability to change. The influence of AI on construction industry will continue expanding beyond what has already been achieved so far[4].

1. **Literature Review:**

For a long time, figuring out how strong concrete is has been important in construction. Usually, people have relied on old-school methods like using formulas based on experience and doing tests in the lab. But now, with the rise of machine learning, we can do things in a smarter way. Machine learning lets us to forecast concrete strength more precisely., which is super useful for building stuff [5].

Back in the day, Individuals once guess how strong concrete was by using formulas they came up with from experience. They also did tests in the lab and used specific mixes for different jobs. But these methods weren't always super accurate. They didn't always consider everything that could be impacted by how strong the concrete ended up being [5]**.**

Using machine learning to predict concrete strength is really promising. Researchers like Akçaözeklı and Ahti’s discovered that employing regression models and neural networks can be good at this. These models examine a wide range of variables, such as the amount of cement utilized. The ratio of water to cement, and how the concrete is cured. Compared to the outdated methods, they can provide more accurate predictions because they take all these factors into account [6].

Recent research, like the studies by Yazdani and colleagues [7], as well as others by Oskoei and team, stress the significance of understanding how various characteristics impact predictions of concrete strength. It truly does matter what kind of cement is used and how the concrete is cured. However, there are still certain difficulties. For example, understanding the complicated Machine learning models can be difficult. Plus, we need bigger sets of data that cover all kinds of construction situations and environmental conditions.

In the future, researchers will keep improving machine learning techniques, using bigger and more diverse sets of data, and testing models in real-life situations. If these obstacles are overcome, machine learning models that forecast concrete strength more correctly may improve. This could change the way we build things and make sure structures are strong and safe.

**Exiting Research Paper:**

Existing research on machine learning regression has explored various algorithms and methodologies.

**Paper-1:** Data-driven model for ternary-blend concrete compressive strength prediction using machine learning approach by Babatunde Abiodun Salami a, Teslim Olayiwola b, Tajudeen A. Oyehan c, Ishaq A. Raji d[8].

Model Summary: The study proposes a machine learning approach to predict the compressive strength of ternary-blend concrete. It highlights the limitations of existing models and introduces an optimized least square support vector machine (LSSVM) using coupled simulated annealing (CSA). The optimized LSSVM-CSA model significantly improves prediction accuracy, outperforming other models with an R² value of 0.954. Sensitivity analysis confirms the model's robustness and strong correlation between input and target variables. Overall, the LSSVM-CSA model offers a superior and reliable alternative for predicting concrete compressive strength.



Result:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | LSSVM-CSA |  | GP |  |
| Performance | Train | Test | Train | Test |
| R² | 0.982 | 0.954 | 0.859 | 0.894 |
| AARD | 4.834 | 7.407 | 16.277 | 14.453 |
| ARD | -1.052 | -0.734 | -2.162 | 0.236 |
| RMSE | 2.230 | 3.335 | 5.918 | 4.662 |
| SD | 0.099 | 0.119 | 0.269 | 0.236 |

Fig summary of model performance

**Paper-2:** Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques by Surya Abisek Rajakarunakaran, Arun Raja Lourdu, Suresh Muthusamy, Hitesh Panchal, Ali Jawad Alrubaie, Mustafa Musa Jaber, Mohammed Hasan Ali, Iskander Tlili, Andino Maseleno, Ali Majdi, Shahul Hameed Masthan Ali [9]

Model Summary: The study examines machine learning techniques to predict the compressive strength of self-compacting concrete (SCC) based on mix design parameters like cement, silica flour, aggregates, polypropylene fiber, water, superplasticizer, and VMA. Among various regression models analysed, random forest regression performs best, achieving the lowest error metrics and highest R². The study highlights the effectiveness of machine learning, particularly random forest regression, in optimizing SCC mix designs for desired compressive strength.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | MSE | MAE | R² |
| Linear Regression | 5.79 | 33.53 | 3.73 | 0.82 |
| Lasso Regression | 6.02 | 36.27 | 3.91 | 0.81 |
| Ridge Regression | 5.77 | 33.31 | 3.80 | 0.82 |
| Decision Tree Regressor | 4.12 | 17.01 | 1.91 | 0.91 |
| Multi-Layer perceptron | 6.05 | 36.64 | 3.97 | 0.80 |

**Paper-3:** Unboxing machine learning models for concrete strength prediction using XAI Sara Elhishi2, Asmaa Mohammed Elashry & Sara El‑Metwally[ 10].

Model Summary: The paper explores machine learning regression for predicting self-compacting concrete (SCC) strength, with random forest regression yielding the highest accuracy. Additionally, convolution-based ensemble learning models are effective predictors. XAI techniques are highlighted for understanding key factors. The framework includes data stages and identifies significant parameters. Notably, XGBoost achieves a high R-square score of 0.91, underscoring its efficacy in SCC strength prediction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | MSE | MAE | R² |
| Liner Regression | 10.28 | 105.76 | 8.23 | 0.57 |
| Lasso Regression | 10.68 | 114.11 | 8.65 | 0.54 |
| Ridge Regression | 10.29 | 105.84 | 8.24 | 0.57 |
| SVM | 9.13 | 83.89 | 7.44 | 0.66 |
| Decision Tree | 6.65 | 44.24 | 4.47 | 0.82 |
| Random Forest | 5.21 | 27.17 | 3.53 | 0.89 |
| XGBoost | 4.37 | 22.33 | 3.04 | 0.91 |

**Paper-4:** Concrete Strength Prediction Using Machine Learning Methods Cat Boost, k‐Nearest Neighbors, Support Vector Regression Alexey N. Beskopylny, Sergey A. Stel’makh, Evgenii M. Shcherban’, Levon R. Mailyan, Besarion Meskhi, Irina Razveeva, Andrei Chernil’nik and Nikita Beskopylny [11]

Model Summary: The paper explores how machine learning can predict mechanical properties of construction materials, like concrete, and improve production processes in construction. It compares CatBoost gradient boosting, k-nearest neighbors, and support vector regression algorithms, finding k-nearest neighbors to be the most promising for production processes and quality control. Highlighting the importance of digitalization, it emphasizes how AI can analyse large datasets from material experiments, advancing construction technology and reducing time costs. The study's novelty lies in establishing relationships between experimental data, models, and machine learning methods. It also reports Cat Boost’s remarkable achievement with an R-squared of 0.91, showing its potential for enhancing predictive accuracy and optimizing construction processes.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MAE | MSE | RMSE | MAPE | R² |
| CatBoost | 2.17 | 7.8 | 2.79 | 6.84 | 0.98 |
| K-nearest Neighbors | 1.97 | 6.85 | 2.62 | 6.15 | 0.99 |
| SVR | 2.61 | 11.39 | 3.37 | 7.89 | 0.98 |

1. **Methodology:**

3.1 Dataset Description:



We gathered dataset from Kaggle. These details refer to different types of concrete in machine learning model that includes the following: water (in liters), sand (in kilos per cubic meter), crushed stone (in kilos per cubic meter), cement (in kilograms per cubic meter), slag (in kilograms per cubic meter), and additives (in kilograms).These features serve as inputs used to forecast the concrete's compressive strength, which is the parameter of interest and is measured in megapascals (MPa). The machine learning models that are applied to the dataset are based on this link between the characteristics and the anticipated parameter [12].

A screenshot of a data

Description automatically generated

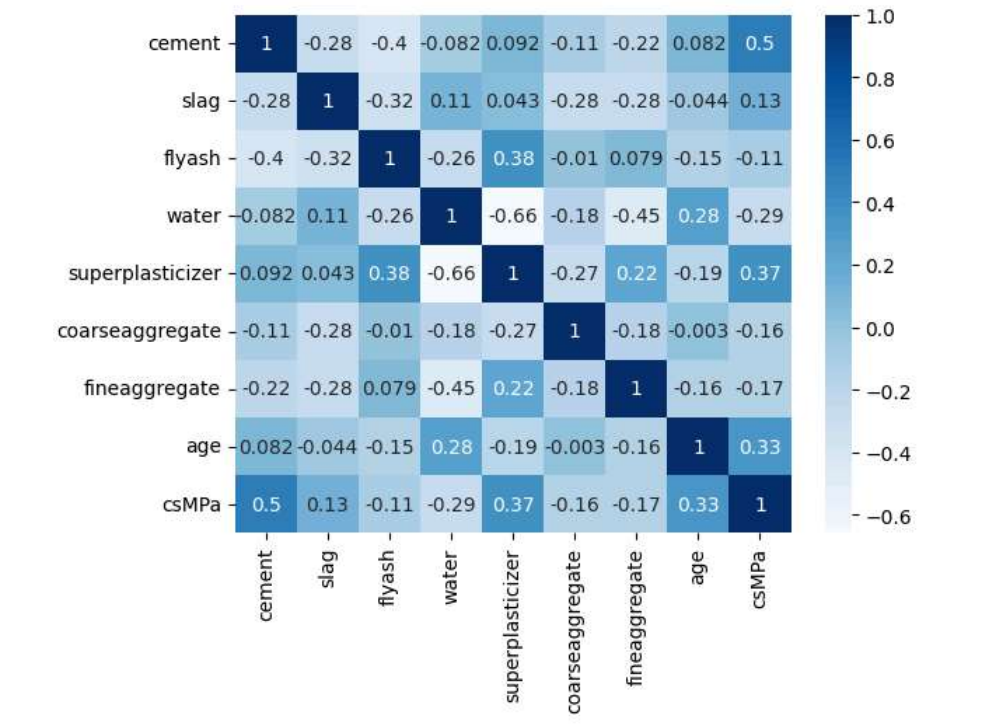
****

Fig. 1 Correlation Matrix

In Figure 1, we can see how the several factors are related to each other. It's evident that each input variable and the output variable have a close relationship, with a correlation of more than 0.5. Additionally, we observe certain instances where when one variable goes up, another goes down, showing a negative correlation [13].

3.2 Data Preprocessing**:**

Some step of data preprocessing

1. Data Cleaning**:** We check our data for any mistakes or missing information. If we find any, we either fix them or remove the problematic data points.
2. Normalization: We make sure that every data set has the same scale. This aids in our model's comprehension of the relative value of each characteristic. We scale features to fit within the same range, for instance, if they measure in liters and kilos.
3. Feature Engineering: Sometimes, the data we have might not be in the best form for our model to understand. So, we create new features or modify existing ones to make them more useful. For instance, we might combine the amount of water and cement to calculate the water-to-cement ratio, which can be a key factor in predicting concrete strength.
4. Outlier Detection and Removal**:** We look for any data points that seem unusual compared to the rest of the dataset. These outliers can mess up our predictions, so we either fix them or remove them.
5. Feature Selection: We figure out which features are the most important for predicting concrete strength. This helps us focus on the most relevant information and avoid unnecessary complexity.

By doing all these steps, we make sure our data is in good shape for the purpose of teaching our machine learning model. This way, we can make moreaccurate predictions about how strong our concrete will be [14].

3.3 Machine learning Model:

3.3.1 Multiple Regression:

Multiple regression is a statistical method used to analyze the relationship between two or more independent variables and a dependent variable. It extends simple linear regression, which deals with just one independent variable[15]. In multiple regression, the relationship between the dependent variable 𝑦*y* and 𝑝*p* independent variables 𝑥1,𝑥2,...,𝑥𝑝*x*1​,*x*2​,...,*xp*​ is represented by the equation:

y=β0​+β1​x1​+β2​x2​+...+βp​xp​+ϵ

where:

* 𝑦*y* is the dependent variable.
* 𝑥1,𝑥2,...,𝑥𝑝*x*1​,*x*2​,...,*xp*​ are the independent variables.
* 𝛽0*β*0​ is the intercept (constant term).
* 𝛽1,𝛽2,...,𝛽𝑝*β*1​,*β*2​,...,*βp*​ are the coefficients (regression coefficients) representing the relationship between each independent variable and the dependent variable.
* 𝜖*ϵ* is the error term, representing the difference between the observed and predicted values of the dependent variable.
  + 1. K-Nearest Neighbour :

K-nearest neighbors (KNN) is a simple, non-parametric machine learning algorithm used for classification and regression. It predicts the output based on the 'k' closest training examples in the feature space[16].

**Distance Metric**: Typically, the Euclidean distance is used to measure the closeness between points.

**Parameter 'k'**: The number of nearest neighbors to consider for making the prediction.

The Euclidean distance between two points 𝑥𝑖*xi*​ and 𝑥𝑗, *xj*​ in an n-dimensional space is given by:

d(*xj,xi*)=2

Steps:

* 1. Compute Distance: Calculate the distance between the new data point and all training data points.
  2. Distances: Sort these distances in ascending order.
  3. Select Neighbors: Select the k-nearest data points.
  4. The output is the average (or weighted average) of the values of the k-nearest neighbors.
     1. Decision Tree :

A decision tree for regression is a technique used to predict a continuous target variable by partitioning the data into subsets based on the input features. Each node in the tree represents a feature, and each branch represents a decision rule or condition. The goal is to split the data in a way that minimizes the variance within each subset[17].

The impurity measure for regression trees is typically the variance or mean squared error (MSE):

MSE=2

* + 1. Random Forest :

Random forests for regression combine the strengths of decision trees and ensemble learning, leading to more accurate and robust predictions. The key lies in the aggregation of multiple uncorrelated trees, which helps to mitigate overfitting and improve generalization[18].

Key Components and Process:

1. Bootstrap Sampling
2. Decision Trees
3. Aggregation

Advantage:

Reduction in Overfitting: Averaging multiple trees reduces the risk of overfitting, which is common in individual decision trees.

Handling Non-linear Relationships: Random forests can model complex interactions and non-linear relationships.

Robustness: They are robust to outliers and noise in the data.

3.4 Model Workflow:

The workflow of the model training and evaluation process as depicted in the flowchart involves several key steps:

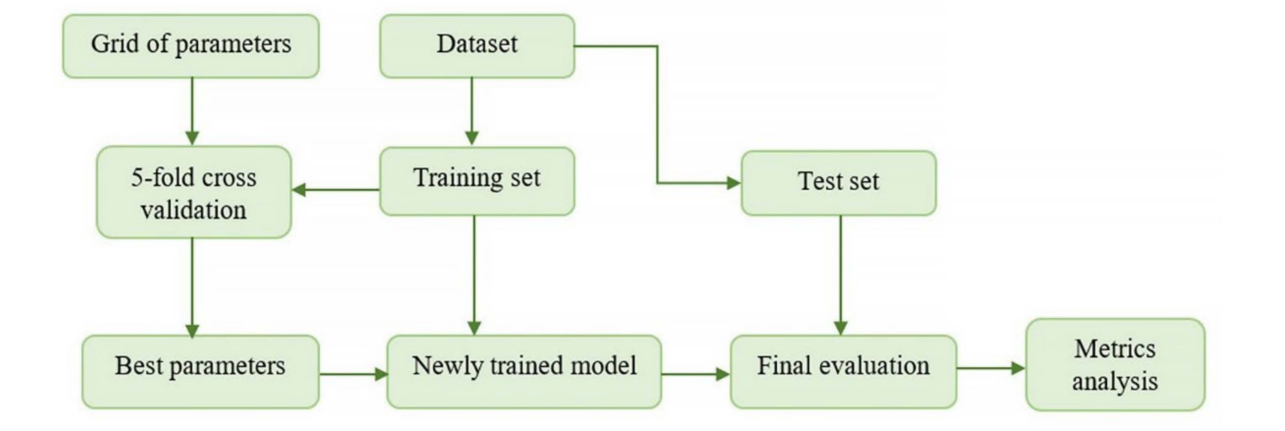


Fig.2 Workflow of Model

1. **Dataset**: The overall dataset is used to create two subsets: the training set and the test set
2. **Grid of Parameters**: This is the initial step where a range of hyperparameters are defined. These hyperparameters are variables that control the learning process and the structure of the model, such as learning rate, depth of trees, number of layers, etc.
3. **5-Fold Cross Validation**: The grid of parameters is passed into a 5-fold cross-validation process. This involves splitting the dataset into five parts, using four parts for training and one part for validation. This process is repeated five times, each time with a different validation set. The performance of the model is evaluated for each combination of hyperparameters, and the average performance across the five folds is calculated.
4. **Best Parameters**: After evaluating the performance of the model for each combination of hyperparameters, the set of parameters that resulted in the best average performance across the folds is selected as the best parameters.
5. **Training Set:** This subset of the data is used to train the model. The training set is used in conjunction with the best parameters identified in the cross-validation step to train the newly trained model.
6. **Newly Trained Model**: With the best parameters obtained from the cross-validation process, a new model is trained on the training set.
7. **Test Set:** This subset of the data is kept separate and is not used during the training process. It is used for the final evaluation of the model to assess how well it generalizes to unseen data.
8. **Final Evaluation:** The newly trained model is evaluated on the test set to assess its performance. This evaluation provides an unbiased estimate of the model’s performance on unseen data.
9. **Metrics Analysis:** Finally, the performance metrics from the final evaluation are analysed. These metrics could include accuracy, precision, recall, F1 score, etc., depending on the specific problem and evaluation criteria.

3.5 Metrics for Machine Learning Models:

When analysing regression models, it is important to use various evaluation metrics to evaluate their performance. This study uses five metrics: mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) of determination and the coefficient R2[19].

Mean Absolute Error:

MAE calculates the average absolute difference between the predicted and actual values. It gives an indication of the average magnitude of errors without considering their direction.

**MAE =**

Mean Squared Error (MSE):

MSE calculates the average of the squared differences between the predicted and actual values. It penalizes larger errors more significantly than MAE, making it sensitive to outliers.

**MSE =2**

Root Mean Squared Error (RMSE):

The average size of mistakes in the same units as the target variable is measured by RMSE, which is the square root of MSE.

**RMSE=**

R-squared (R2):

R-squared quantifies the percentage of the target variable's variance that the regression model accounts for. Higher values indicate a better fit, and the range is 0 to 1, where yˉ is the mean of the observed values.

**R2=**

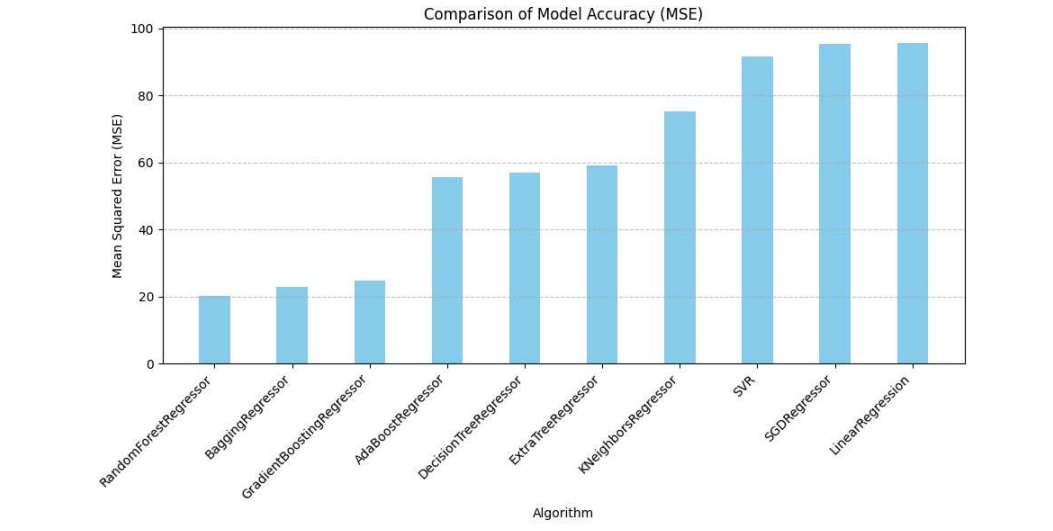


Fig.2 Comparison of Accuracy

1. **Experimental setup:**

In Python, pandas serve as the cornerstone for data manipulation and analysis, facilitating tasks such as reading datasets, preprocessing, and managing Data Frames, while seaborn and matplotlib excel in crafting visually appealing and insightful statistical graphics. Complementing these visualization tools, scikit-learn empowers users with a robust suite of machine learning capabilities, spanning data preprocessing, modelling, and evaluation, including regression algorithms. Leveraging Python's simplicity, versatility, and extensive support in terms of machine learning and data science, these libraries thrive in extracting insights from datasets. With 8GB of RAM and ample storage capacity, this arsenal of tools is poised to tackle data processing, model training, and evaluation with efficiency and effectiveness [20].

To refine the Random Forest model's performance, we follow a methodical process. Initially, we choose the best initial model based on our evaluation criteria. Then, we pinpoint which aspects of the model to adjust, like the number of trees or their maximum depth. Next, we use GridSearchCV to systematically test various combinations of these adjustments, searching for the setup that performs best. We validate each setup using a reliable method like k-fold cross-validation to ensure our results are solid. Once we find the optimal settings, we apply them to our model and assess its performance on a separate test dataset. This helps us gauge how well our tuned model will perform on new, unseen data. By following these steps, we fine-tune our Random Forest model to achieve the best possible results across different situations [21].

The K-fold pass-validation technique is a commonly employed strategy. The dataset is divided into k same-sized folds. The model is trained ok times, whenever the remaining fold is used for validation and okay-1 folds are used for training. This manner is repeated k instances, with every fold serving because the validation set exactly once. The very last performance metric is commonly the common of the overall performance metrics acquired across all folds [22].

1. **Result:**

The real values from the dataset are shown next to the expected values produced by our model in Table 2's prediction error plot. This visual aid allows us to see how much the real and predicted values differ from one another, giving us information on the model's accuracy and performance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | MSE | MAE | RMSE | R2 |
| Multiple Regression | 95.617 | 7.864 | 9.778 | 0.637 |
| K Nearest Neighbour | 75.134 | 6.883 | 8.668 | 0.714 |
| Decision Tree | 57.128 | 4.857 | 7.558 | 0.783 |
| Random Forest | 21.733 | 3.279 | 4.661 | 0.917 |

Table 2 Matrix develop by model.

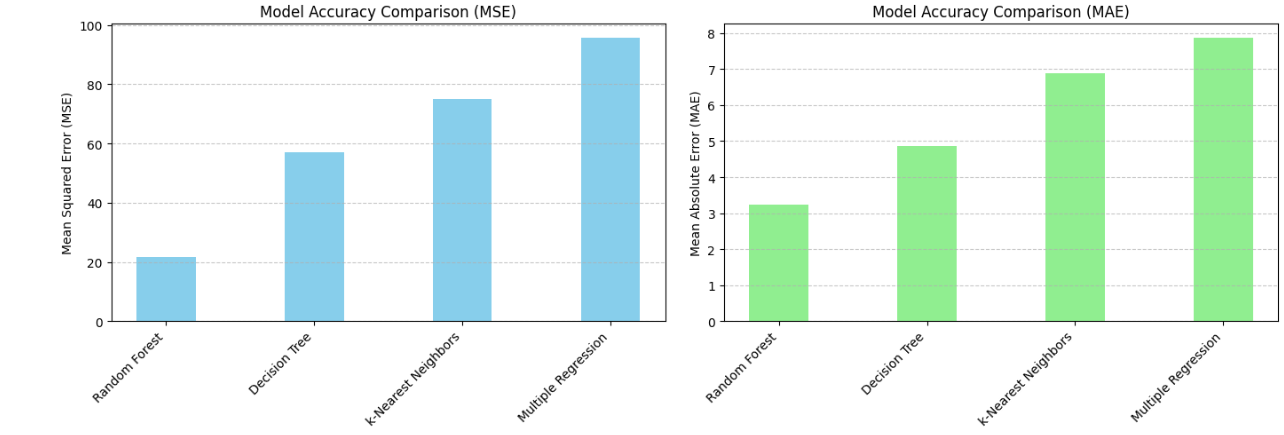
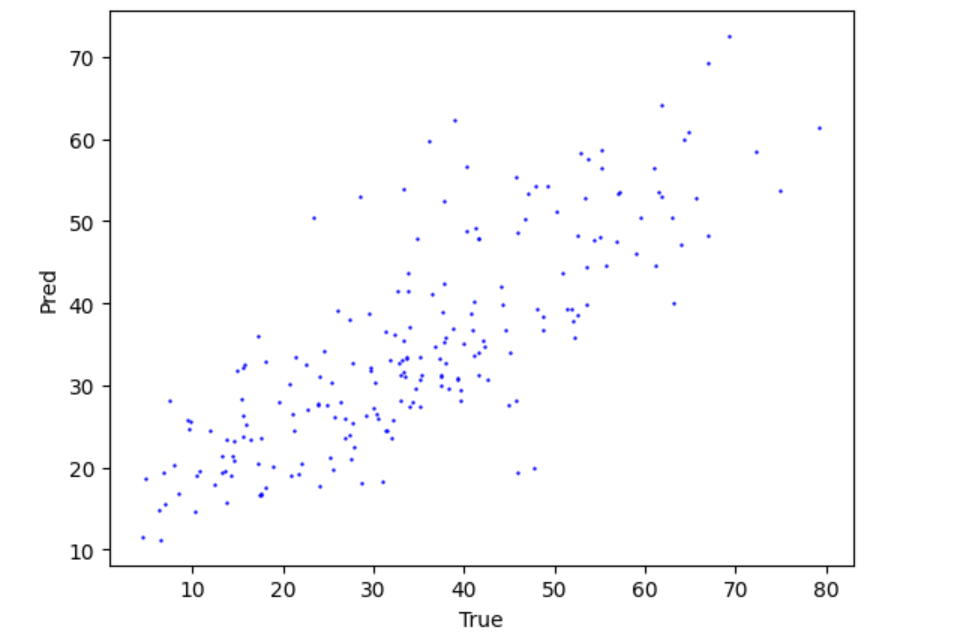


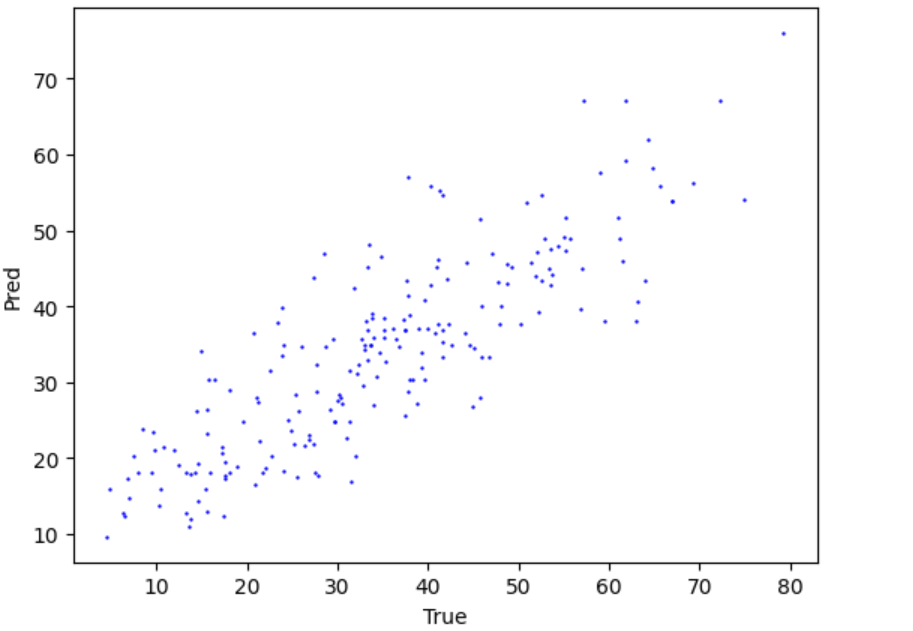
Fig. . Metric values for the developed regression models: (a) MAE; (b) MSE

A graph of different colored bars

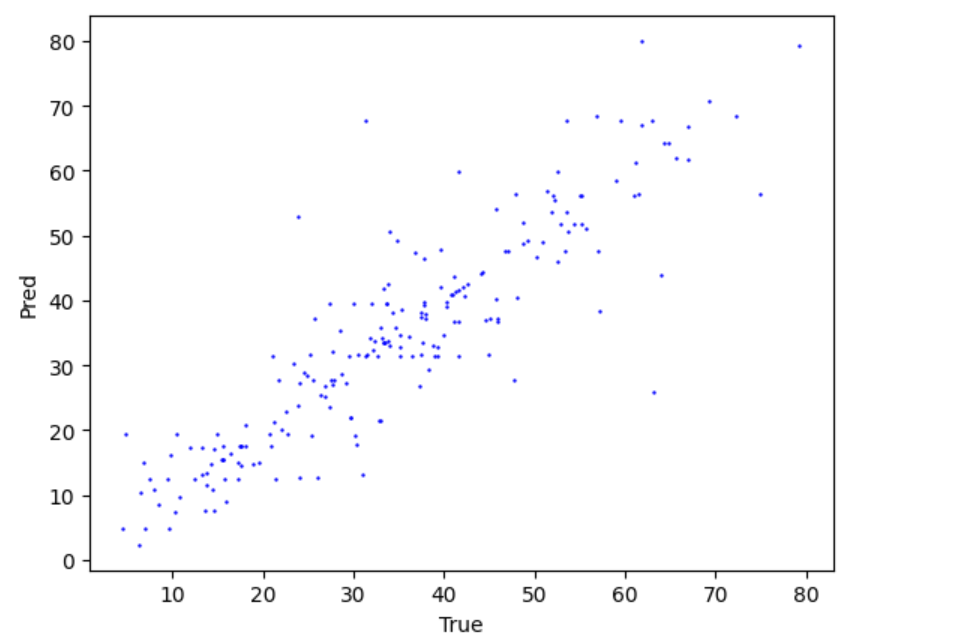
Description automatically generated with medium confidence Fig.3 Metric values for the developed regression models: (a) RMSE; (b) R2



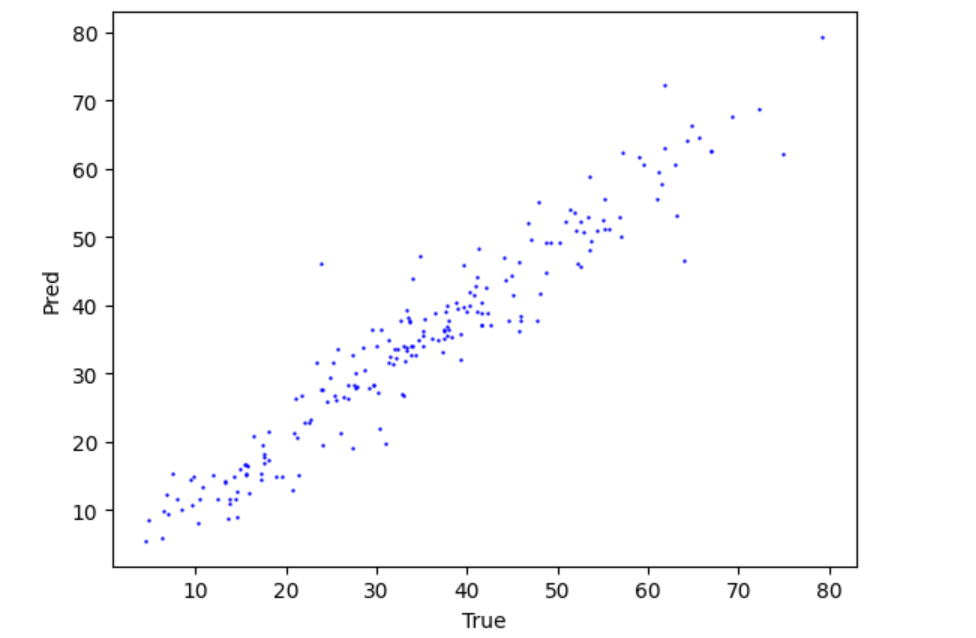
(a)



(b)



(c)



(d)

Fig.3 8. Relationship between actual compressive strength and calculated values (a)

Linear Regression (b)K-Nearest Neighbour (c) Decision Tree (d) Random Forest.

1. **DISCUSSION:**

When we use machine learning on data from testing concrete, it's crucial to remember that Concrete is a complicated material influenced by numerous variables. These factors can cause concrete's characteristics to vary, even if we don't fully understand or control them all. So, there will always be some level of inconsistency or error in the data, usually within about 10%. This level of variation is considered normal and acceptable in the concrete production industry. Essentially, it means that no matter how careful we are, there will always be some unpredictability when working with concrete data [23].

The flaws that we detected in our investigation, including the Root Mean Squared Error (RMSE) of 4.6 to 9.7, the Mean Absolute Error (MAE) of 3.2 to 7.8, and the Mean Squared Error (MSE) of 21.6 to 95.6, are in line with the findings of other researchers. This indicates that our findings are consistent with those published by other writers who have previously looked at issues related to these.

1. **Conclusion:**

Three distinct machine learning algorithms—multiple regression, k-nearest neighbours (KNN), and decision tree random forest—were examined in this study. These methods allowed us to forecast the self-compacting concrete's compressive strength. To do this, we gathered a lot of data from various sources and applied it to our models. The goal was to see which algorithm worked best for this specific task.

Artificial intelligence approaches have been found to be an accurate means of predicting the compressive strength of self-compacting concrete. The mean absolute percentage error (MAPE) of our models ranged from 3.15% to 7.89%. This indicates that, on average, the errors between our forecasts and the actual compressive strength values were relatively small, usually ranging from 3.15% to 7.89%.

The random forest method yielded the most accurate predictions out of the four machine learning algorithms we examined. Its coefficient of determination (R-squared) was the highest, and its mistakes were the least. There were four different types of errors: the mean squared error (MSE) was 21.7, the root mean squared error (RMSE) was 4.6, the mean absolute error (MAE) was 3.2, and the R-squared value was 0.917. In comparison to the other algorithms examined, the random forest method had the highest accuracy and minimum errors in its ability to forecast the compressive strength of self-compacting concrete.

The methods we developed can be effectively used in the production and quality control of building materials. They don't need extensive computing resources, making them practical for real-world applications. Additionally, in the future, we could create an expert artificial intelligence-based intelligence. This system would gather all the experimental data we've collected and store it in an electronic format accessible at universities. It could then provide valuable information to workers and researchers interested in advancing the industry.

The models we've developed can be thoroughly checked and approved for forecasting self-compacting concrete's compressive strength. This assessment considers all the data we have available.

**Reference:**

|  |  |
| --- | --- |
| [1] | A. M. Neville, Properties of concrete, Longman Publishers. |
| [2] | chandra.S.Bjornstron, Influence of superplasticizer type and dosage on the slump loss of portland cement mortars, 2002. |
| [3] | D. D. B. Divekar-1030, Ferrocement Construction Manual, Shivaji Nagar, Model Colony, Pune.  [4] https://doi.org/10.1016/j.conbuildmat.2016.03.214 |

[5]Chandra, S.; Björnström, J. Influence of superplasticizer type and dosage on the slump loss of Portland cement mortars—Part II. Cem. Concr. Res. 2002, 32, 1613–1619. <https://doi.org/10.1016/S0008‐8846(02)00838‐4>.

[6] Farooq, F.; Czarnecki, S.; Niewiadomski, P.; Aslam, F.; Alabduljabbar, H.; Ostrowski, K.A.; Śliwa‐Wieczorek, K.; Nowobilski, T.; Malazdrewicz, S. A Comparative Study for the Prediction of the Compressive Strength of Self‐Compacting Concrete Modi‐ fied with Fly Ash. Materials 2021, 14, 4934. <https://doi.org/10.3390/ma14174934>.

[7]//scholar.google.com/citations?view\_op=view\_citation&hl=en&user=ZaTDvKwAAAAJ&citation\_for\_view=ZaTDvKwAAAAJ:zYLM7Y9cAGgC

[8] Data-driven model for ternary-blend concrete compressive strength prediction using machine learning approach by Babatunde Abiodun Salami a, Teslim Olayiwola b, Tajudeen A. Oyehan c, Ishaq A. Raji d

[9]Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques by Surya Abisek Rajakarunakaran, Arun Raja Lourdu, Suresh Muthusamy, Hitesh Panchal, Ali Jawad Alrubaie, Mustafa Musa Jaber, Mohammed Hasan Ali, Iskander Tlili, Andino Maseleno, Ali Majdi, Shahul Hameed Masthan Ali

[10] Unboxing machine learning models for concrete strength prediction using XAI Sara Elhishi2, Asmaa Mohammed Elashry & Sara El‑Metwally.

[11] Concrete Strength Prediction Using Machine Learning Methods Cat Boost, k‐Nearest Neighbors, Support Vector Regression Alexey N. Beskopylny, Sergey A. Stel’makh, Evgenii M. Shcherban’, Levon R. Mailyan, Besarion Meskhi, Irina Razveeva, Andrei Chernil’nik and Nikita Beskopylny

[12] <https://www.kaggle.com/datasets/maajdl/yeh-concret-data>

[13] Understanding Correlation Matrices by [Alexandria Hadd](https://us.sagepub.com/en-us/nam/author/alexandria-r-hadd) ,[Joseph Lee Rodgers](https://us.sagepub.com/en-us/nam/author/joseph-l-rodgers).

[14] [Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work](https://amzn.to/3b5yutA),” 2012

[15] Multiple Regression: A Primer by Paul D. Allison:

[16] Lectures on the Nearest Neighbor Method by Gérard Biau and Luc Devroye (Springer Series in the Data Sciences):

[17] https://www.cambridge.org/core/books/understanding-machine-learning/decision-trees/F1C6923D0974184C4F161B406AD1393C

[18] Random Forests with R by Robin Genuer and Jean-Michel Poggi

[15] Random Forests with R by springer (Robin Genure).

[16]<https://www.researchgate.net/publication/311797168_Prediction_of_concrete_mix_strength_using_random_forest_model>.

[17]<https://www.researchgate.net/publication/294560839_A_Review_on_Self_Compacting_Concrete>

[18] **Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlowby** Aurélien Géron

[19**] An Introduction to Statistical Learning** by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani

[20] **Applied Predictive Modeling** by Andreas Gelman, John B. Carlin, Hal S. Stern, David B. Rubin, Miriam Weinstein, Donald B. Rubin

[21] Beskopylny, A.N.; Shcherban’, E.M.; Stel’makh, S.A.; Mailyan, L.R.; Meskhi, B.; Evtushenko, A.; Varavka, V.; Beskopylny, N. Nano‐Modified Vibrocentrifuged Concrete with Granulated Blast Slag: The Relationship between Mechanical Properties and Micro‐Structural Analysis. Materials 2022, 15, 4254. <https://doi.org/10.3390/ma15124254>.

[22]Beskopylny, A.N.; Stel’makh, S.A.; Shcherban’, E.M.; Mailyan, L.R.; Meskhi, B.; Varavka, V.; Beskopylny, N.; El’shaeva, D. A Study on the Cement Gel Formation Process during the Creation of Nanomodified High‐Performance Concrete Based on Nanosilica. Gels 2022, 8, 346. <https://doi.org/10.3390/gels8060346>

[23][. Hundreds page of machine learning](https://www.google.com/search?sca_esv=de93c5c4281f23a7&cs=1&sxsrf=ACQVn080GYZ1pG4Zp7qT-sZiJP7PWezsUg:1714672024788&q=The+Hundred-page+Machine+Learning+Book&stick=H4sIAAAAAAAAAE1Uv2_TQBSOKzUkbis1qShSxBC1EqqKUGyffwUhtQKKGBoh0SxMVs6_Lv5x9tmHY3tEYq8Y-AOYGJgZGBjYqBBIDB1RxcbUBSEWJFyIL73t3ffe-77vPftay1trA3cgikgAvpMJRNoMJyaaYrsf2JMET7Hbh1Hkpyfc_7w0V2kBdb2O67o6NjXZCYwQn3CrF7EEQxUZ1KthqBHR0DBl6amgDgVYXsJVOPOqcn4QDgTZ0KSZMu-FgToDqgAZtTgUw1Au57lSjhLAdMFZnJqJp9Y6ZE8OU1erYSIjaAiyyGRix4UZYrJi3ykQYbJdB0IDLaYQED_xCofZ8AJP0YHMcComfuFKDqs38jIr0mxOJ4LEC0TLYk5UKwZFuRgajgXLZd0dUmKiefFiaAos3IyZCRyimuLQuzRzShfiK1gmWazWsZUTM9ccibEDPUhiP2F0YkIBhQYzJ1OISspWDCkpYhDkLD9UhjMgFQu5uidLgMlFw4yKBDO-1CvzEhgKy1cyS01ltjpXtoqZLi72bPoOSAnrjxyVAF1h06OJJZphnFzaZYGMhRyZKEYWwxpWDKTowYypMYdSUBWw7pqDLaozd3FGYgpE5t73zFBBhcnYEhQHbvmdO-NW1n_8_rbRO-VevPl4yn3h-PXDKErtoHhsBxNqW-Ooe41vHmA6pUV3rbfCty--WxFVarsH_MqRTcfRKLKmTtFVuzLfHtkhtJP0kdPd5vl7URDYJp1GuHu1t8F3Bia7GPz7Q28v7SxtPZEevPv68nNzv9OozqfnR_u9nd0O37wfhZMp7ug_r5xdPz_f293gW-NJHuEoLDqjzT_u9p23e1vtdlWzf_NXhTeOl7hn719_aLZa3HpDWmo1ysbqq-UbY2T3Hz7FVmJbt-KJa_dH89fisH4t7lZajpvcX7EmswxZBAAA&sa=X&ved=2ahUKEwip9cCHw--FAxXHTGwGHbsiBoMQ7fAIegUIABDNBQ)

[24].[Machine learning for absolute beginner : A plain English Introduction](https://www.google.com/search?sca_esv=de93c5c4281f23a7&cs=1&sxsrf=ACQVn080GYZ1pG4Zp7qT-sZiJP7PWezsUg:1714672024788&q=Machine+Learning+for+Absolute+Beginners:+A+Plain+English+Introduction&stick=H4sIAAAAAAAAAFWUv2_TQBTH40gNidtKTSqKFDFY7VKVIfFvp0JqCwQJqRGIdmGycv51ts_nX1fH9ojEXjHwBzAxMDMwMLBRIZAYOqKKjakLQixIuFV9Kd6ev--97-e9s6-9sL48cAY8n-YKKYCmCWvB1IAutjhkTRPsYocDYeinJ8z_eXUMh6JvZ8O4jg1VspEe4BNm6SIWQKBAnXi1DNSY11VMaHo6VEZDUF7TFTDzqnJ2EAyGkq4KM_mqFxaVmagMAbXmR3wQSOVVrpDDRKRcYBalRuIpNYfkSUHqqLUcSxDoQ4mnmNh2QAYpVuTbBYwptmMDoMP5FlDsJ15h0zE85MmaKFGd8IlfOIJN6_W8zIo0u7LjxcRDvGnSSRQzEotyvjQcDU2HdrfjEseqF82XJoPCyegwyI4Vgx9513ZOyBy-kqU4i5Q6NvPYyFVboO6ihpLIT6gdnxCRAJ0OJxEAS0KPGJC4iESU0_xAHs1EoZjjap4kiBQXjjLCx5j6pV6Zl6Iu03w5M5VUokfnSGYx0_j5ORu-LaYx7Q9tJRY1mW6PJCZvBFFy7SwLqM9xpFjWswjUsqxDWUMzSmOMBFQV0O6qjU2i0emiLI6IyNPpfc8IZFgY1C2BEXLKH8wZs7jy88_31f4p8_Ltp1PmK8Ou7IdhaqHiqYWmxDIPw94ttjXGxCVFb7m_yHYuvlseVrS9Mbt4YJHDcBKarl30lJ7EdiZWAKwkfWz3Nlj2foiQZRA3xL2b_VW2OzDoi8HlH7rd3GyuPxMevv_26ktrt9uons8vDnb7m1tdtvUgDKYu7mq_bpzdPj_f2Vpl24fTPMRhUHQna3-djbvvdtY7napm987vSm8cN5nnH958bLXbzEpDaLYbZWPp9cJ4cnU57NeXgx0m3B5IQ3RELO6e5bgYV8jb3B73BFWO3Bg7yE0h9wiTJDSPLnmPW8w_ffAHr3gEAAA&sa=X&ved=2ahUKEwip9cCHw--FAxXHTGwGHbsiBoMQ7fAIegUIABC9BQ)

[25][. Wikipedia](https://en.wikipedia.org/wiki/Machine_learning)

[26]//scholar.google.com/citations?view\_op=view\_citation&hl=en&user=ZaTDvKwAAAAJ&citation\_for\_view=ZaTDvKwAAAAJ:zYLM7Y9cAGgC

[27]https://scholar.google.co.in/scholar?q=Oskouei+concrete+strength+research+paper&hl=en&as\_sdt=0&as\_vis=1&oi=scholart